

Use of Bayesian statistics to improve optical measurement uncertainty by combined multi-tool metrology

Nien Fan Zhang¹, Richard M. Silver², Hui Zhou², Bryan M. Barnes²

¹ Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD

² Mechanical Metrology Division, National Institute of Standards and Technology, Gaithersburg, MD

Abstract

Recently, there has been significant research investigating new optical technologies for dimensional metrology of features 32 nm in critical dimension and smaller. When modeling optical measurements a library of curves is assembled through the simulation of a multi-dimensional parameter space. A nonlinear regression routine described in this paper is then used to identify the optimum set of parameters that yields the closest experiment-to-theory agreement. However, parametric correlation, measurement noise, and model inaccuracy all lead to measurement uncertainty in the fitting process for optical critical dimension (OCD) measurements. To improve the optical measurements, other techniques such as atomic force microscopy (AFM) and scanning electronic microscopy (SEM) can also be used to provide supplemental *a priori* information. In this paper, a Bayesian statistical approach is proposed to allow the combination of different measurement techniques that are based on different physical measurement. The effect of this approach will be shown to reduce the uncertainties of the parameter estimators.

Key Words: Best linear unbiased estimator; generalized least squares; nonlinear regression; optical critical dimension measurements; scatterfield microscopy.

Introduction

Recently, there has been significant research investigating new optical technologies for dimensional metrology of features 32 nm in critical dimension and smaller. Among them are scatterometry and more recently scatterfield microscopy, a technique that combines well-defined angle-resolved illumination with image-forming optics [1][2][3]. Experimental results using these optical techniques have demonstrated nanometer accuracy across a range of target dimensions as small as 40 nm with pitch values 1/10 the measurement wavelength [4]. To achieve these results using the angle-resolved scatterfield technique, reflected intensities are measured using a well-characterized optical microscope operated in a scanned illumination mode. In addition, electromagnetic scattering simulations are generated using idealized geometric representations of the measured features. These representations are designed so that the quantitative parameters that define the dimensions and material properties of the idealized structures correspond to the physical attributes of interest, such as line width and line height. Comprehensive libraries of simulated reflectivity curves are generated using well-developed rigorous coupled waveguide or finite difference time domain electromagnetic scattering models [5][6][7][8] with input parameters that cover the actual physical values. Quantitative critical dimension measurements can then be achieved through a comprehensive parametric analysis, in which the experimental signature is compared against the

simulation library. This method of parametric fitting is often used in quantitative scatterometry. Not only can a set of parameters be found that best describes the measured object, but the parametric fitting process can also yield quantitative critical dimension measurements with measurement uncertainties that include all instrumentation and sample components that contribute to the measurement uncertainty [9]. In the work presented here, we develop a method to embed *a priori* measurement information in the regression fitting algorithm and achieve reduced measurement uncertainties.

When modeling optical measurements a library of curves is assembled through the simulation of a multi-dimensional parameter space. A nonlinear regression routine is then used to identify the optimum set of parameters that yields the closest experiment-to-theory agreement. This approach assumes that the model is adequately describing the physical conditions and that an acceptable goodness-of-fit is achieved with the best set of parameters. However, parametric correlation, measurement noise, and model inaccuracy all lead to measurement uncertainty in the fitting process for optical critical dimension (OCD) measurements as shown in [9]. To improve the optical measurements, other techniques may be used to augment the parametric fitting. As an example reference measurements for critical dimensions can assist in the definition of the central starting point and the boundaries of the multi-dimensional parameter space. Reference metrology techniques such as atomic force microscopy (AFM) and scanning electronic microscopy (SEM) that are used to provide supplemental *a priori* information can also be used to validate the library fitting results. Historically, in the statistical regression analysis optical metrology simulation and experimental data are treated independently from the

supplemental reference metrology data obtained from AFM or SEM, for example. In this paper we establish rigorous statistical methods to combine measurement information from different metrology sources directly in the parametric fitting process and thereby improve the measurement uncertainty.

The Bayesian statistical approach proposed in this paper allows the combination of different measurement techniques that are based on different physical measurement principles (*e.g.*, probe scanning, electron scanning, *etc.*) that often have varying sensitivities to physical properties such as edge runout and surface roughness. Additionally, this approach allows the inclusion of measurements of the sample optical properties from a different hardware platform (*e.g.* an ellipsometer) and rigorously incorporates these measurements and their uncertainties in the regression algorithm. Including independently obtained information on the optical constants, feature geometry, or any other attribute of the measurement system, influences the best fit values and their associated uncertainties. In Section 2, we discuss the use of nonlinear regression models in scatterfield microscopy for OCD analysis. In Section 3, a Bayesian statistical approach is introduced to use *a priori* information of the parameters and thus to obtain optimal posterior parameter estimators and their corresponding uncertainties. The effect of this approach will be shown to reduce the uncertainties of the parameter estimators. As a detailed illustration, two physical examples are presented in Section 4 followed by conclusions.

2. Nonlinear regression models for OCD study

The technique of scatterfield microscopy has been described elsewhere in detail [1][2][3][10]. A complete set of scatterfield microscope measurements includes y_1, \dots, y_N , which are the measured values of an interested variable Y , *e.g.*, intensity, and x_1, \dots, x_N , which represent the measurement conditions, *e.g.*, the values of wavelength or angle, under which the N data points y_1, \dots, y_N are obtained correspondingly. As mentioned in the Introduction, electromagnetic scattering simulations can be performed at each of x_1, \dots, x_N based upon a representation of the sample defined using K measurement/tool parameters. The simulated optical response is denoted by $y(x_i; \mathbf{a}), i = 1, \dots, N$, where $\mathbf{a} = \{a_1, \dots, a_K\}^T$ is a parameter vector representing the floating (*i.e.* variable) parameters, for example, line height, line width, line edge roughness (LER), and sidewall height, *etc.* Our goal is to compare $\{y_1, \dots, y_N\}$ with $\{y(x_i; \mathbf{a})\}, i = 1, \dots, N$, the simulated values under the condition of $x_i, i = 1, \dots, N$ for the parameters $\mathbf{a} = \{a_1, \dots, a_K\}^T$, to find an optimal estimator of the parameter vector \mathbf{a} .

One illustration of \mathbf{a} , used also for Section 4, is shown below as Figure 1. Here, five lines show five parameters that completely define a geometric model of the sample. The dotted lines represent the top, middle, and bottom line widths. The solid lines indicate the pitch, p , and the line height, h , of this periodic structure. Any number of these parameters may be floated, with the remainder defined by fixed values. Additional floating parameters may be included to improve the model, such as the optical constants,

n and k , of the material. The individual floating parameters used in the model are the components of the vector \mathbf{a} .

In general $y(x_i; \mathbf{a})$ is a nonlinear function of the parameter vector \mathbf{a} . Treating $y(x_i; \mathbf{a})$ as a mean response of y_i , we have a nonlinear regression for y_i and $y(x_i, \mathbf{a})$ for $i = 1, \dots, N$ given by

$$y_i = y(x_i, \mathbf{a}) + \varepsilon_i \quad \text{for } i = 1, \dots, N, \quad (1)$$

where ε_i is the corresponding random error with zero mean. Using a first order Taylor expansion at a specific point of the vector \mathbf{a} (e.g., an initial value or an optimal value), $\mathbf{a}(0) = \{a_1(0), \dots, a_K(0)\}^T$, a linear approximation of that nonlinear function is given by

$$y(x_i; \mathbf{a}) \sim y(x_i; \mathbf{a}(0)) + \sum_{k=1}^K \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)} (a_k - a_k(0)), \quad (2)$$

where $y(x_i; \mathbf{a}(0))$ is the simulated value of $y(x_i; \mathbf{a})$ at $\mathbf{a}(0)$. From (1) we have an approximation of the nonlinear regression model given by

$$y_i = y(x_i; \mathbf{a}(0)) + \sum_{k=1}^K \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)} (a_k - a_k(0)) + \varepsilon_i, \quad i = 1, \dots, N, \quad (3)$$

where $\left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)}$ is the value of the partial derivative of $y(x_i; \mathbf{a})$ with respect to a_k

at $\mathbf{a} = \mathbf{a}(0)$. See [11] p. 473. The covariance matrix of $\boldsymbol{\varepsilon} = \{\varepsilon_1, \dots, \varepsilon_N\}^T$ is denoted by \mathbf{V} . In this paper, we assume that the random variables $\{\varepsilon_i\}$ are uncorrelated. Namely, \mathbf{V} is a diagonal matrix denoted by $\mathbf{V} = \text{diag}[\sigma_1^2, \dots, \sigma_N^2]$. In general, \mathbf{V} does not have to be a diagonal matrix. By re-parameterization, the linear model in (3) is expressed as

$$y_i(0) = \sum_{k=1}^K D_{ik}(0) \beta_k(0) + \varepsilon_i, \quad i = 1, \dots, N, \quad (4)$$

where $\beta_k(0) = a_k - a_k(0)$, (5)

$$D_{ik}(0) = \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)}, \quad (6)$$

and

$$y_i(0) = y_i - y(x_i; \mathbf{a}(0)). \quad (7)$$

In (4), $\beta_k(0)$ for $k=1, \dots, K$ are the regression parameters. $D_{ik}(0)$ and $y_i(0)$ are the values for the explanatory or predictor variables and the response variable of the regression model, respectively. A matrix form of (4) is given by

$$\mathbf{Y}(0) = \mathbf{D}(0) \cdot \boldsymbol{\beta}(0) + \boldsymbol{\varepsilon}, \quad (8)$$

where $\mathbf{Y}(0) = (y_1(0), \dots, y_N(0))^T$ is a N by 1 vector and

$$\mathbf{D}(\mathbf{0}) = \begin{bmatrix} D_{11}(0) \dots D_{1K}(0) \\ \dots \\ D_{N1}(0) \dots D_{NK}(0) \end{bmatrix} \quad (9)$$

is a N by K matrix. $\boldsymbol{\beta}(\mathbf{0}) = (\beta_1(0), \dots, \beta_K(0))^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$. The mean of $\boldsymbol{\varepsilon}$ is $\mathbf{0}$ and the covariance matrix of $\mathbf{Y}(\mathbf{0})$ and $\boldsymbol{\varepsilon}$ is \mathbf{V} . The best linear unbiased estimator (B.L.U.E.) of $\boldsymbol{\beta}(\mathbf{0})$ is the generalized least squares (GLS) estimator given by

$$\hat{\boldsymbol{\beta}}(\mathbf{0}) = \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}). \quad (10)$$

See [12], p. 123. Namely, among all the linear unbiased estimators of $\boldsymbol{\beta}(\mathbf{0})$, $\hat{\boldsymbol{\beta}}(\mathbf{0})$ is the one with the smallest variance. See *e.g.*, [13], p. 88. Here, a linear estimator of $\boldsymbol{\beta}(\mathbf{0})$ based on (8) is a linear combination of $y_1(0), \dots, y_N(0)$. In the case that \mathbf{V} is a diagonal matrix, the model in (8) is a weighted least squares regression. See [13], p. 164. From (5), the B.L.U.E. of $\mathbf{a} = \{a_k; k = 1, \dots, K\}$ is $\hat{\mathbf{a}}$ given by

$$\hat{a}_k = \hat{\beta}_k(\mathbf{0}) + a_k(\mathbf{0}), \quad (11)$$

for $k = 1, \dots, K$. An example of best-fit data are shown in Figure. 2. The covariance matrix of $\hat{\mathbf{a}} = \{\hat{a}_1, \dots, \hat{a}_K\}^T$ is given by

$$\text{Cov}[\hat{\mathbf{a}}] = \text{Cov}[\hat{\boldsymbol{\beta}}(\mathbf{0})] = \left(\mathbf{D}(\mathbf{0})^T \cdot \mathbf{V}^{-1} \cdot \mathbf{D}(\mathbf{0}) \right)^{-1}. \quad (12)$$

See [12], p. 124. The standard deviations or standard uncertainties of $\hat{\beta}_k(0)$ or \hat{a}_k , $k = 1, \dots, K$ are given by the square roots of the diagonal elements of $\text{Cov}[\hat{\mathbf{a}}]$ and denoted by $\sigma_{\hat{a}_k}$. Note that when $\{\varepsilon_i\}$ are Gaussian distributed, from (10) and (11) $\hat{\beta}(0)$ and $\hat{\mathbf{a}}$ are also Gaussian distributed. Note also that we can use the Gauss-Newton method [14], p. 40 – 41 to iteratively improve $\hat{\mathbf{a}}$ and keep improving the estimates until there is no change. Thus, we can assume that $\mathbf{a}(0) = \{a_1(0), \dots, a_K(0)\}^T$ in (2) is an estimate vector from the Gauss-Newton method.

3. Bayesian statistical analysis and the use of prior information of parameters

Recent studies have shown that OCD measurements are fundamentally limited by correlation of the fitting parameters [15]. Although it is possible that the simulated library of curves and the experimental data are in good agreement, the uncertainty in the fitting process may be greater than desired due to measurement noise and correlation between the fitting parameters. However, quantitative information regarding these parameters, either from other measurement techniques or *a priori* manufacturing knowledge of material parameters such as their optical constants, n and k , may be available and used to improve measurement uncertainties. For example, for the parameter of middle height, we may use another nondestructive tool, such as AFM to get an estimate of middle height with its own uncertainty. We can treat this information based on AFM measurements as prior information for that parameter and apply a Bayesian statistical approach.

In Bayesian analysis, model parameters such as $\boldsymbol{\beta}(0)$ in (8) or equivalently, $\{a_k\}$, $k=1,\dots,K$ are treated as random and have their own probability distributions. In particular, we assume that among the K parameters, the first p ($p \leq K$) parameters, a_1, \dots, a_p have prior information with their means given by

$$E[\mathbf{a}] = \mathbf{a}^* = (a_1^*, \dots, a_p^*)^T \quad (13)$$

and a known covariance matrix. From (5), the means of the corresponding adjusted regression parameters are given by

$$E[\beta_k(0)] = a_k^* - a_k(0) \quad (14)$$

and denoted by $\beta_k^*(0)$ for $k=1,\dots,p$. We assume that $\{a_i; i=1,\dots,p\}$ are uncorrelated from each other and the covariance matrix of $(a_1, \dots, a_p)^T$ or equivalently the sub-vector of $\boldsymbol{\beta}(0)$ denoted by $\boldsymbol{\beta}_p(0) = (\beta_1(0), \dots, \beta_p(0))^T$ is given by

$$\Sigma_{\boldsymbol{\beta}_p(0)} = \begin{pmatrix} \sigma_{a_1}^2, 0, 0, \dots, 0 \\ 0, \sigma_{a_2}^2, 0, \dots, 0 \\ \dots\dots\dots \\ 0, 0, \dots\dots\dots, \sigma_{a_p}^2 \end{pmatrix}. \quad (15)$$

In general, $\{a_i\}$ can be correlated leading to a non-diagonal covariance matrix in (15). Referring to the regression model in (8), from [15], p. 382 – 384, we treat the prior information on $\boldsymbol{\beta}_p(0)$ as p additional “data points” of the response variable in (8). For

example, for the first parameter $\beta_1(0)$, we assume that it has a prior distribution with a mean of $\beta_1^*(0) = a_1^* - a_1(0)$ and a variance of $\sigma_{a_1}^2$. Considered as a function of $\beta_1(0)$, the prior distribution can be viewed as an “observation” $\beta_1^*(0)$ with all “explanatory variables” in the regression model equal to zero except the first one. Namely, an additional regression equation similar to (8) is given by

$$\beta_1^*(0) = \beta_1(0) + \varepsilon_{N+1}, \quad (16)$$

where $E[\varepsilon_{N+1}] = 0$ and $\text{Var}[\varepsilon_{N+1}] = \sigma_{a_1}^2$. In general, for $k = 1, 2, \dots, p$, we have

$$\beta_k^*(0) = \beta_k(0) + \varepsilon_{N+k} \quad (17)$$

where

$$\beta_k^*(0) = a_k^* - a_k(0). \quad (18)$$

Combining (16) and (17) with (8), we have an expanded linear model given by

$$\mathbf{Y}^*(\mathbf{0}) = \mathbf{D}^*(\mathbf{0}) \cdot \boldsymbol{\beta}(\mathbf{0}) + \boldsymbol{\varepsilon}^*, \quad (19)$$

where the $(N + p)$ by K matrix

$$\mathbf{D}^*(\mathbf{0}) = \begin{pmatrix} \mathbf{D}(\mathbf{0}) \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} D_{11}(0), \dots, D_{1K}(0) \\ \dots\dots\dots \\ D_{N1}(0), \dots, D_{NK}(0) \\ 1, 0, 0, \dots, \dots, 0 \\ 0, 1, 0, \dots, \dots, 0 \\ \dots\dots\dots \\ 0, 0, 0, \dots, 0, 1, 0, \dots, 0 \end{pmatrix} \quad (20)$$

with $\mathbf{1}$ a $p \times K$ matrix consisting of p row vectors of length of K with only a single 1 in each row and the other elements in each row are zeros as shown in the second equality.

In (18),

$$\mathbf{Y}^*(\mathbf{0}) = (y_1(0), \dots, y_N(0), \beta_1^*(0), \dots, \beta_p^*(0))^T \quad (21)$$

and

$$\boldsymbol{\varepsilon}^* = (\varepsilon_1, \dots, \varepsilon_N, \varepsilon_{N+1}, \dots, \varepsilon_{N+p})^T, \quad (22)$$

with $E[\boldsymbol{\varepsilon}^*] = \mathbf{0}$ and the covariance matrix of $\boldsymbol{\varepsilon}^*$ given by

$$\mathbf{V}^* = \text{Cov}[\boldsymbol{\varepsilon}^*] = \text{diag}[\sigma_1^2, \dots, \sigma_N^2, \sigma_{a_1}^2, \dots, \sigma_{a_p}^2]. \quad (23)$$

Similar to (10), the posterior estimators of $\boldsymbol{\beta}(0)$ based on the GLS are given by

$$\hat{\boldsymbol{\beta}}^\#(0) = (\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}))^{-1} \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{Y}^*(\mathbf{0}) \quad (24)$$

with the posterior covariance matrix of the parameter estimators given by

$$\text{Cov}[\hat{\mathbf{a}}^\#] = \text{Cov}[\hat{\boldsymbol{\beta}}^\#(\mathbf{0})] = \left(\mathbf{D}^*(\mathbf{0})^T \cdot \mathbf{V}^{*-1} \cdot \mathbf{D}^*(\mathbf{0}) \right)^{-1}, \quad (25)$$

where $\hat{\mathbf{a}}^\# = \hat{\boldsymbol{\beta}}^\#(\mathbf{0}) + \mathbf{a}(\mathbf{0})$ are the posterior estimators of the original parameters \mathbf{a} . It is clear that $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is the B.L.U.E. of $\boldsymbol{\beta}(\mathbf{0})$ based on the expanded model in (19). Namely, under the model in (19) among all the linear unbiased estimators of $\boldsymbol{\beta}(\mathbf{0})$, $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is the one with the smallest variance. Here, a linear estimator of $\boldsymbol{\beta}(\mathbf{0})$ based on (19) is a linear combination of $y_1(\mathbf{0}), \dots, y_N(\mathbf{0}), \beta_1^*(\mathbf{0}), \dots, \beta_p^*(\mathbf{0})$. Note that both $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ and $\hat{\boldsymbol{\beta}}(\mathbf{0})$ are unbiased estimators of $\boldsymbol{\beta}(\mathbf{0})$. The standard uncertainties of the posterior estimators $\hat{a}_k^\#$, $k = 1, \dots, K$ are given by the square roots of the diagonal elements of $\text{Cov}[\hat{\mathbf{a}}^\#]$ and denoted by $\sigma_{\hat{a}_k^\#}$.

Following that $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ in (24) is the B. L. U. E. of $\boldsymbol{\beta}(\mathbf{0})$, it is demonstrated in Appendix 2 that $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is also a linear function of $\hat{\boldsymbol{\beta}}(\mathbf{0})$ and $\boldsymbol{\beta}^*(\mathbf{0})$ as well as a function of the variances and covariances of these estimators. In particular, when $K = 2$ and $p = 1$,

$$\hat{\beta}^\#(\mathbf{0}) = \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \hat{\beta}(\mathbf{0}) + |\mathcal{Q}_1| \frac{\beta_1^*(\mathbf{0})}{\sigma_{a_1}^2} \left(\begin{array}{c} \text{Var}[\hat{\beta}_1(\mathbf{0})] \\ \text{Cov}[\hat{\beta}_1(\mathbf{0}), \hat{\beta}_2(\mathbf{0})] \end{array} \right) + \left(\begin{array}{c} 0 \\ \frac{g_2}{\sigma_{a_1}^2} \end{array} \right) \right], \quad (26)$$

where $|\mathcal{Q}|$ and $|\mathcal{Q}_1|$ are the determinants of $\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0})$ in (25) and $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0})$ in (12), respectively, and g_2 are the second components of $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0})$ (see A.19). From (26), it is obvious that when $\sigma_{a_1} = \infty$, i.e., there is no prior information for the first

parameter, and $\hat{\beta}^\#(0) = \hat{\beta}(0)$. Note that when the random errors, $\{\varepsilon_i\}$ are Gaussian distributed, from (24) and (11) $\hat{\beta}^\#(0)$ and $\hat{\mathbf{a}}^\#$ are also Gaussian distributed.

We investigate the posterior covariance matrix in (25). By expanding it, we have

$$\begin{aligned}
\mathbf{D}^*(\mathbf{0})^T \cdot \mathbf{V}^{*-1} \cdot \mathbf{D}^*(\mathbf{0}) &= (\mathbf{D}(\mathbf{0})^T, \mathbf{1}^T) \begin{pmatrix} \mathbf{V}, \mathbf{0} \\ \mathbf{0}, \Sigma_{\beta_p(0)} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{D}(\mathbf{0}) \\ \mathbf{1} \end{pmatrix} \\
&= (\mathbf{D}(\mathbf{0})^T, \mathbf{1}^T) \begin{pmatrix} \mathbf{V}^{-1}, \mathbf{0} \\ \mathbf{0}, \Sigma_{\beta_p(0)}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{D}(\mathbf{0}) \\ \mathbf{1} \end{pmatrix} \\
&= \mathbf{D}(\mathbf{0})^T \cdot \mathbf{V}^{-1} \cdot \mathbf{D}(\mathbf{0}) + \mathbf{1}^T \Sigma_{\beta_p(0)}^{-1} \mathbf{1}.
\end{aligned} \tag{27}$$

In (27),

$$\mathbf{1}^T \Sigma_{\beta_p(0)}^{-1} \mathbf{1} = \begin{pmatrix} 1/\sigma_{a_1}^2, \dots, 0, \dots, 0 \\ \dots \\ 0, \dots, 1/\sigma_{a_p}^2, \dots, 0 \\ 0, \dots, 0 \\ \dots \\ 0, \dots, 0 \end{pmatrix} \equiv \begin{pmatrix} \Sigma_{\beta_p(0)}^{-1}, \mathbf{0} \\ \mathbf{0}, \mathbf{0} \end{pmatrix}, \tag{28}$$

where $\mathbf{0}$ indicates a zero matrix with an appropriate size. We compare the variances of the estimators given by (10) and the variances of the posterior estimators given by (25).

In Appendix 1, it is shown that

$$\text{Var}[\hat{\beta}_k^\#(0)] \leq \text{Var}[\hat{\beta}_k(0)] \tag{29}$$

for $k = 1, \dots, K$. Or equivalently, $\text{Var}[\hat{a}_k^\#] \leq \text{Var}[\hat{a}_k]$ for $k = 1, \dots, K$. This indicates that the variance of a posterior parameter estimator is equal to or smaller than that of the

corresponding usual GLS estimator without prior information of the model parameters. In addition, no matter which parameter the prior information is used for, the variances of all posterior estimators are equal to or smaller than those of the corresponding usual GLS estimators. The two variances are the same if and only if $\sigma_{a_i}^2 = \infty$ for $i = 1, \dots, p$, *i.e.*, if there is no prior information for all these p model parameters. Note that since both estimators are unbiased estimators of the model parameters, the posterior estimators have smaller mean squared errors than those of the corresponding usual GLS estimators, correspondingly.

In addition, we also compare the variances of the posterior estimators with the prior variances of the model parameters. In Appendix 1, it is shown that

$$\text{Var}[\hat{\beta}_i^\#(0)] \leq \sigma_{\beta_i}^2 \quad (30)$$

for $i = 1, \dots, p$. Thus, the posterior variances are smaller than the prior variances of the model parameters, correspondingly.

From (29) and (30), it is clear that when we use prior information about the regression model parameters from other metrology sources to OCD study, the resultant uncertainties of the posterior estimators are smaller than both the prior uncertainties and the uncertainties of the regular GLS estimators of the model parameters. Therefore, by using the Bayesian analysis to combine measurement results from multiple sources, the resultant uncertainties are improved.

4. Two examples

To illustrate the methodology, two sets of measurements were performed on etched features on well-characterized wafers. A first set of experiments involved the measurement of line arrays on a silicon-on-silicon wafer produced by SEMATECH* using the OMAG-3 reticle set [17]. There are multiple nearly identical arrays on the wafer, each nominally 100 nm in line width with a 200 nm space, yielding a 300 nm pitch and thus are labeled on the wafer as “L100P300” arrays. The actual line widths vary among arrays located throughout the wafer due to pattern variations induced by various combinations of lithographic exposure dose and focus across the wafer in a focus-exposure matrix. As illustrated in Figure 2, measurements were conducted by performing two orthogonal scans for one polarization, and then repeated for a second polarization, yielding four scans for each array. Measurements were made at 17 angles for each scan with four scans per target (two polarizations at each of the two orthogonal scan angles) resulting in $N = 68$ for each experimental data set. As indicated in Figure 2, for this example, the variable Y is the reflectivity and the variable X represent the measurement conditions, the angle and the corresponding polarization and scan direction. In Equation 3, y_i ($i = 1, \dots, 68$) are the measured reflectivity and $y(x_i; a)$ ($i = 1, \dots, 68$) are the simulated values of reflectivity. In the simulation, there are three floating parameters ($K = 3$), i.e., top, middle, and bottom line widths as shown in Figure. 1. The matrix for the explanatory variables in the regression model, $\mathbf{D}(\mathbf{0})$ in (8) and (9) is a 68×3 matrix calculated by (6) based on simulations and spline interpolations. The $\mathbf{Y}(\mathbf{0})$ from (8) is a vector of size of 68. The covariance matrix of $\boldsymbol{\varepsilon}$, \mathbf{V} is a 68×68 diagonal matrix based on

measurements. The original parameters $\mathbf{a}(0) = (121, 115, 141)^T$ are a set of optimal values for top, middle, and bottom line widths. From (10), the GLS of the adjusted parameters are given by $\hat{\boldsymbol{\beta}}(0) = (-1.32, -3.68, 2.24)^T$ and from (11) the original parameters $\hat{\mathbf{a}}(0) = (119.68, 111.32, 143.24)^T$ for top, middle, and bottom line widths. The corresponding standard deviations, *i.e.*, standard uncertainties are 0.84, 2.23, and 1.32, as shown in Table 1. In this example, quantitative reference information from AFM metrology is available. Table 1 lists the mean values of the three parameters given by $\mathbf{a}^* = (119.21, 117.32, 132.87)^T$ with the corresponding standard uncertainties equal to 0.75 in the corresponding parentheses from AFM metrology. In this example we only use the means and standard uncertainties for the parameters of top and middle line widths as prior information and thus, treat these two pieces of information as two data points in (18), *i.e.*, $p = 2$. Now the covariance matrix \mathbf{V}^* in (23) is a 70 x 70 diagonal matrix with the last two diagonal elements equal to 0.75. From (24), the Bayesian estimators of the three adjusted and the original parameters are $\hat{\boldsymbol{\beta}}^\#(0) = (0.27, 1.03, -0.54)^T$ and corresponding $\hat{\mathbf{a}}^\# = (121.27, 116.03, 140.46)^T$ with the corresponding standard uncertainties $\sigma_{\hat{\mathbf{a}}^\#} = (0.30, 0.68, 0.44)^T$. It is obvious that the uncertainties of $\hat{\boldsymbol{\beta}}^\#(0)$ are also smaller than the prior uncertainty of 0.75. The results show a change in the estimates of the parameters as well as an improvement in the uncertainties. The resulting uncertainties from the combined measurements are smaller than each of the individual measurement results.

A similar second set of experiments was completed for line patterns etched into a nitride film on a polysilicon substrate. These gratings were measured in the same manner as before, although here 21 angles per scan were measured. The four combinations of polarizations and scan axes yield an experiment for which $N = 84$. Notably, the simulation library was expanded to five floating parameters ($K = 5$). These nitride patterns were parameterized by again using the top, middle, and bottom line widths from Figure. 1. In addition, the height of the line and the optical constant n of the nitride film were also allowed to vary. The optical constant variable n is expressed here as a percentage of a nominal value provided *a priori* but without an uncertainty. Reference metrology data were also acquired using an AFM, yielding the height and the top, middle, and bottom widths.

Both the Type A and Type B uncertainties were evaluated based on the optical reflectivity data. The Type A uncertainty was evaluated from the standard deviation of the mean reflectivity at each data point and can be approximated as 5 % of the measured reflectivity. The Type B uncertainties were evaluated through parametric uncertainty analysis and estimations of the spectral width of the light-emitting diode (LED), the choice of simulation model, the reflectivity changes from adding 1 nm of line-edge roughness (LER), and the calibration of the incident angle. Model choice and spectral width contributed insignificantly to the Type B estimate, while the largest component was the possibility of LER, which also is approximately 5% of the reflectivity. The uncertainty from angular calibration contributes to the Type B as a function of increasing angle. Between cross-correlation among floating parameters and the combined

uncertainties related to the reflectivity measurements, the parametric uncertainties for the OCD fitting results for this example are large, as seen in the OCD fitting column of Table 2.

Comparisons were made between the measured data and the simulation library with and without the AFM reference data. Calculations were carried out in the same way as for the previous example. Table 2 shows the parametric results and the measured AFM values as well as their corresponding standard uncertainties in the parentheses.

Embedding the four AFM parameters into the parametric fitting affects each of the best-fit parametric values and their uncertainties differently. For example, the standard uncertainty for the top line width from the embedded metrology is just slightly better than that provided from AFM. However, the standard uncertainty for the bottom line width has decreased by almost a factor of two with respect to that for the AFM measurement. The embedded AFM reference values allow the standard uncertainty for the optical constant n for the nitride film to be halved with respect to the OCD measurement. The goodness-of-fit of this embedded metrology parameterization for OCD is shown as Figure 3. The goodness of fit data are quite good with small residuals.

5. Conclusions

In this paper, a Bayesian statistical approach has been applied to combine measurement information from other reference metrology platforms into OCD regression analysis. The

resultant estimators of the model parameters have smaller variances and smaller mean squared errors than those based on the measurements from optical measurements alone. The measurement uncertainties are also improved. The new methodology has important implications in devising measurement strategies that take advantage of the best measurement attributes of each individual technique. This method may be applied to other metrology methods such as model-based scanning electron microscopy or quantitative ellipsometry.

A key result of this approach is that a combination of measurements may be conceived that optimizes measurement uncertainty for a given measurement throughput. For example by combining film thickness measurements and optical constants measurements on one platform with OCD measurements and known manufacturing variations, a substantial gain in measurement uncertainty can be achieved with simultaneous improvements in throughput. A second important possibility is improved calibrations and measurement uncertainties. This approach has immediate appeal to combine measurement techniques that achieve the lowest overall uncertainty for a given structure. In the midst of sub - 20 nm sized features of increasing geometric complexity, this method will allow the best rigorous combination of metrology platforms, each best suited to a particular aspect of the measurement, resulting in the most complete and lowest uncertainties.

References

- [1] Silver R M, Barnes B M, Attota R, Jun R, Stocker M, Marx E, and Patrick H J 2007 Scatterfield microscopy for extending the limits of image-based optical metrology, *Applied Optics* **46**, 4248-4257
- [2] Patrick H J, Attota R, Barnes B M, Germer T A, Dixson R G, Stocker M T, Silver R M, and Bishop M R 2008 Optical critical dimension measurement of silicon grating targets using back focal plane scatterfield microscopy, *Journal of Micro/Nanolithography, MEMS and MOEMS* **7**, 013012-11
- [3] Barnes B M, Attota R, Howard L P, Lipscomb P, Stocker M T, and Silver R M 2007 Zero-Order and Super-Resolved Imaging of Arrayed Nanoscale Lines using Scatterfield Microscopy, *AIP Conference Proceedings* **931**, 397-401
- [4] Silver R M, Barnes B M, Zhou H, Zhang N F, and Dixson R 2009 Angle-resolved optical metrology using multi-technique nested uncertainties, *Modeling Aspects in Optical Metrology II, Munich, Germany: Proc. SPIE* **7390**, 73900P-12 (2009).
- [5] Moharam M G, Grann E B, Pommet D A, and Gaylord T K 1995 Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings, *J. Opt. Soc. Am. A* **12**, 1068–1076
- [6] Moharam M G, Pommet D A, Grann E B, and Gaylord T K 1995 Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach, *J. Opt. Soc. Am. A* **12**, 1077–1086
- [7] Lalanne P and Morris G M Highly improved convergence of the coupled-wave method for TM polarization, *J. Opt. Soc. Am. A* **13**, 779–784

- [8] Taflove A 1975 Numerical-solution of steady-state electromagnetic scattering problems using time-dependent Maxwell's Equations, *IEEE Trans. Microwave Theory Tech.* **23**, 623-630
- [9] Silver R M, Zhang N F, Barnes B M, Zhou H, Heckert A, Dixson R, Germer T A, and Bunday B 2009 Improving optical measurement accuracy using multi-technique nested uncertainties, *Metrology, Inspection, and Process Control for Microlithography XXIII*, San Jose, CA, USA: Proc. SPIE **7272**, 727202-14
- [10] Silver R M, Barnes B M, Heckert A, Attota A, Dixson R, and Jun J 2008 Angle resolved optical metrology, *Metrology, Inspection, and Process Control for Microlithography XXII*, San Jose, CA, USA: Proc. SPIE **6922**, 69221M-12
- [11] Neter J, Wasserman W, and Kutner M H 1983 *Applied Linear Regression Models*, Richard D. Irwin, Homewood, Illinois
- [12] Ravishanker N and Dey D K 2002 *A first Course in Linear Model Theory*, Chapman & Hall/CRC, Boca Raton, Florida
- [13] Stapleton J H 1995 *Linear Statistical Models*, Wiley, New York
- [14] Bates D M and Watts D G 1988 *Nonlinear Regression Analysis and Its Applications*, Wiley, New York
- [15] Silver R, Germer T A, Attota R, Barnes B M, Bunday B, Allgair J, Marx E, and Jun J 2007 Fundamental limits of optical critical dimension metrology: a simulation study, *Metrology, Inspection, and Process Control for Microlithography XXI*, San Jose, CA, USA: SPIE, Proc. SPIE **6518**, 65180U-17
- [16] Gelman A, Carlin J B, Stern H S, and Rubin D B 2004 *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, Florida

- [17] Bunday B, Bishop M, and Silver R 2003 The OMAG3 Reticle Set, 2003
SEMATECH Tech Transfer Document 03074417A-ENG
- [18] Golub G H and Van Loan C H 1989 *Matrix Computations*, The Johns Hopkins
University press, Baltimore
- [19] Rao C R and Rao M B 1998 *Matrix Algebra and its Applications to Statistics and
Econometrics*, World Scientific, Singapore.

Appendix 1: Proof of Equations 29 and 30.

First we consider the case of $K = 2$. Since the cases of $p = 1$ can be treated as a special case of $p = 2$ by letting the corresponding $\sigma_{a_i}^2 = \infty$, we will discuss the case of $p = 2$ only. We denote the symmetric matrix $\mathbf{D}^T(\mathbf{0})\mathbf{V}^{-1}\mathbf{D}(\mathbf{0})$ in (11) by \mathcal{Q}_1 , *i.e.*,

$$\mathcal{Q}_1 = \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) = \begin{pmatrix} q_{11}, q_{12} \\ q_{12}, q_{22} \end{pmatrix}. \quad (\text{A.1})$$

Assuming that $\mathbf{D}(\mathbf{0})$ has a full column rank, $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0})$ is a positive definite matrix because \mathbf{V} is positive definite by Theorem 4.2.1, p. 140 from [18]. The determinant of \mathcal{Q}_1 is given by $|\mathcal{Q}_1| = q_{11}q_{22} - q_{12}^2 > 0$. By Corollary 4.2.2, p. 140 in [18], $q_{11} > 0$, $q_{22} > 0$.

The corresponding term in (25) denoted by \mathcal{Q} is expressed as

$$\begin{aligned} \mathcal{Q} &= \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \\ &= \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) + \mathbf{1}^T \Sigma_{\beta_2(\mathbf{0})}^{-1} \mathbf{1}. \end{aligned} \quad (\text{A.2})$$

$$= \begin{pmatrix} q_{11} + \frac{1}{\sigma_{a_1}^2}, q_{12} \\ q_{12}, q_{22} + \frac{1}{\sigma_{a_2}^2} \end{pmatrix}$$

The determinant of \mathcal{Q} is given by

$$|\mathcal{Q}| = |\mathcal{Q}_1| + \frac{q_{11}}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2} > 0. \quad (\text{A.3})$$

From (25) and (A.3), the variance of the posterior estimator of the first parameter is given by

$$\text{Var}[\hat{\beta}_1^\#(0)] = \frac{q_{22} + \frac{1}{\sigma_{a_2}^2}}{\left|Q_1\right| + \frac{q_{11}}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2}} \quad (\text{A.4})$$

while from (12) and (A.1)

$$\text{Var}[\hat{\beta}_1(0)] = \frac{q_{22}}{\left|Q_1\right|}.$$

The difference between the two variances is given by

$$\begin{aligned} & \text{Var}[\hat{\beta}_1(0)] - \text{Var}[\hat{\beta}_1^\#(0)] \\ &= \frac{\frac{q_{22}^2}{\sigma_{a_1}^2} + \frac{q_{12}^2}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2 \sigma_{a_2}^2}}{\left|Q_1\right| \left|Q\right|} > 0 \end{aligned}$$

Similarly, the result holds for $\beta_2(0)$. Thus, the variance of a posterior parameter estimator is smaller than that of the usual GLS estimator without prior information of model parameters.

If $\mathbf{D}(\mathbf{0})$ does not have a full column rank,

$$\begin{aligned} & \text{Var}[\hat{\beta}_1(0)] - \text{Var}[\hat{\beta}_1^\#(0)] \\ &= \frac{\frac{q_{22}^2}{\sigma_{a_1}^2} + \frac{q_{12}^2}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2 \sigma_{a_2}^2}}{\left|Q_1\right| \left|Q\right|} \geq 0 \end{aligned}$$

For (30), from (A.4) the difference of the prior variances of the model parameter, e.g., $i = 1$ and the corresponding posterior variance is given by

$$\sigma_{a_1}^2 - \text{Var}[\hat{\beta}_1^\#(0)] = \frac{\sigma_{a_1}^2 |Q_1| + \frac{q_{11} \sigma_{a_1}^2}{\sigma_{a_2}^2}}{|Q_1| + \frac{q_{11}}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2}} \geq 0.$$

Now we consider a case with three parameters, i.e., $K = 3$. Since the cases of $p < 3$ can be treated as special cases of $p = 3$ by letting the corresponding $\sigma_{\beta_i}^2 = \infty$, we will discuss the case of $p = 3$.

We denote the symmetric matrix $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0})$ in (12) by Q_1 , i.e.,

$$Q_1 = \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) = \begin{pmatrix} q_{11}, q_{12}, q_{13} \\ q_{12}, q_{22}, q_{23} \\ q_{13}, q_{23}, q_{33} \end{pmatrix}. \quad (\text{A.5})$$

Assuming that $\mathbf{D}(\mathbf{0})$ has a full column rank, $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0})$ is a positive definite matrix because \mathbf{V} is positive definite by Theorem 4.2.1, p. 140 from [18]. The corresponding term in (25) denoted by Q is expressed as

$$\begin{aligned} Q &= \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \\ &= \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) + \mathbf{1}^T \Sigma_{\beta_3(0)}^{-1} \mathbf{1}. \end{aligned} \quad (\text{A.6})$$

$$= \begin{pmatrix} q_{11} + \frac{1}{\sigma_{a_1}^2}, q_{12}, q_{13} \\ q_{12}, q_{22} + \frac{1}{\sigma_{a_2}^2}, q_{23} \\ q_{13}, q_{23}, q_{33} + \frac{1}{\sigma_{a_3}^2} \end{pmatrix}$$

The determinant of Q can be expressed as

$$|Q| = |Q_1| + \frac{M_{11}}{\sigma_{a_1}^2} + \frac{M_{22}}{\sigma_{a_2}^2} + \frac{M_{33}}{\sigma_{a_3}^2} + \frac{q_{33}}{\sigma_{a_1}^2 \sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2 \sigma_{a_3}^2} + \frac{q_{11}}{\sigma_{a_2}^2 \sigma_{a_3}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2 \sigma_{a_3}^2}. \quad (\text{A.7})$$

where the determinants,

$$M_{11} = \begin{vmatrix} q_{22} & q_{23} \\ q_{23} & q_{33} \end{vmatrix} \quad M_{22} = \begin{vmatrix} q_{11} & q_{13} \\ q_{13} & q_{33} \end{vmatrix} \quad \text{and} \quad M_{33} = \begin{vmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{vmatrix}$$

are the (1,1), (2,2), and (3,3) minors of matrix Q_1 . See [19], p. 143. Without loss of generality, we only need to check the variances for the first parameter. From (24) and (A.7), the variance of the posterior estimator of the first parameter is given by

$$\text{Var}[\hat{\beta}_1^\#(0)] = \frac{M_{11} + \frac{q_{33}}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_3}^2} + \frac{1}{\sigma_{a_2}^2 \sigma_{a_3}^2}}{|Q_1| + \frac{M_{11}}{\sigma_{a_1}^2} + \frac{M_{22}}{\sigma_{a_2}^2} + \frac{M_{33}}{\sigma_{a_3}^2} + \frac{q_{33}}{\sigma_{a_1}^2 \sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2 \sigma_{a_3}^2} + \frac{q_{11}}{\sigma_{a_2}^2 \sigma_{a_3}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2 \sigma_{a_3}^2}} \quad (\text{A.8})$$

while from (12) and (A.5)

$$\text{Var}[\hat{\beta}_1(0)] = \frac{M_{11}}{|Q_1|}.$$

In (A.8), the matrices of the corresponding M_{11} , M_{22} , and M_{33} are positive definite (or semidefinite positive when $\mathbf{D}(\mathbf{0})$ does not have a full column rank) and $q_{11} > 0$, $q_{22} > 0$, $q_{33} > 0$ by Corollary 4.2.2, p. 140 in [18]. The difference between the two variances is given by

$$\begin{aligned} & \text{Var}[\hat{\beta}_1(0)] - \text{Var}[\hat{\beta}_1^\#(0)] \\ &= \frac{\frac{M_{11}^2}{\sigma_{a_1}^2} + \frac{M_{11}M_{22} - q_{33}|Q_1|}{\sigma_{a_2}^2} + \frac{M_{11}M_{33} - q_{22}|Q_1|}{\sigma_{a_3}^2} + \frac{q_{33}M_{11}}{\sigma_{a_1}^2 \sigma_{a_2}^2} + \frac{q_{22}M_{11}}{\sigma_{a_1}^2 \sigma_{a_3}^2} + \frac{q_{11}M_{11} - |Q_1|}{\sigma_{a_2}^2 \sigma_{a_3}^2} + \frac{1}{\sigma_{\beta_1}^2 \sigma_{\beta_2}^2 \sigma_{\beta_3}^2}}{|Q_1||Q|} \end{aligned} \quad (\text{A.9})$$

It can be shown that

$$M_{11}M_{22} - q_{33}|Q_1| = (q_{23}q_{13} - q_{12}q_{33})^2 = M_{12}^2 \geq 0. \quad (\text{A.10})$$

Similarly,

$$M_{11}M_{33} - q_{22}|Q_1| = (q_{12}q_{23} - q_{13}q_{22})^2 = M_{31}^2 \geq 0. \quad (\text{A.11})$$

In (A.9),

$$\begin{aligned} q_{11}M_{11} - |Q_1| &= q_{12}M_{12} - q_{13}M_{13} \\ &= q_{12}^2q_{33} - 2q_{12}q_{13}q_{23} + q_{13}^2q_{22}. \end{aligned}$$

We show this term is also non-negative. We first consider the case that $q_{12}q_{13} \geq 0$. In this case,

$$\begin{aligned} & q_{12}^2q_{33} - 2q_{12}q_{13}q_{23} + q_{13}^2q_{22} \\ &= q_{12}^2q_{33} - 2\sqrt{q_{33}q_{22}}q_{13}q_{23} + 2\sqrt{q_{33}q_{22}}q_{13}q_{23} - 2q_{12}q_{13}q_{23} + q_{13}^2q_{22} \\ &= \left(q_{12}\sqrt{q_{33}} - q_{13}\sqrt{q_{22}}\right)^2 + 2q_{12}q_{13}\left(\sqrt{q_{22}q_{33}} - q_{23}\right) \\ &\geq 0 \end{aligned} \quad (\text{A.12})$$

The inequality holds because $M_{11} = q_{22}q_{33} - q_{23}^2 > 0$ and $q_{22}, q_{33} > 0$ leading to

$\sqrt{q_{22}q_{33}} > |q_{23}|$. When $q_{12}q_{13} < 0$, corresponding to (A.12),

$$\begin{aligned} & q_{12}^2q_{33} - 2q_{12}q_{13}q_{23} + q_{13}^2q_{22} \\ &= q_{12}^2q_{33} + 2q_{23}|q_{12}q_{13}| + q_{13}^2q_{22} \\ &= q_{12}^2q_{33} - 2\sqrt{q_{33}q_{22}}|q_{12}q_{13}| + 2\sqrt{q_{33}q_{22}}|q_{12}q_{13}| + 2q_{23}|q_{12}q_{13}| + q_{13}^2q_{22} \\ &= \left(|q_{12}|\sqrt{q_{33}} - |q_{13}|\sqrt{q_{22}}\right)^2 + 2|q_{12}q_{13}|\left(\sqrt{q_{22}q_{33}} + q_{23}\right) \\ &\geq 0 \end{aligned} \quad (\text{A.13})$$

From (A.9) – (A.13), $\text{Var}[\hat{\beta}_1(0)] - \text{Var}[\hat{\beta}_1^\#(0)] \geq 0$.

The general case for $K > 3$ follows in the foot-steps of the case of $K = 3$.

For (29), from (A.8) the difference of the prior variances of the model parameter, e.g., $k = 1$ and the corresponding posterior variance is given by

$$\sigma_{a_1}^2 - \text{Var}[\hat{\beta}_1^\#(0)] = \frac{|Q_1| M_{11} + \frac{M_{22} \sigma_{a_1}^2}{\sigma_{a_2}^2} + \frac{M_{33} \sigma_{a_1}^2}{\sigma_{a_3}^2} + \frac{q_{11} \sigma_{a_1}^2}{\sigma_{a_2}^2 \sigma_{a_3}^2}}{|Q_1| + \frac{M_{11}}{\sigma_{a_1}^2} + \frac{M_{22}}{\sigma_{a_2}^2} + \frac{M_{33}}{\sigma_{a_3}^2} + \frac{q_{33}}{\sigma_{a_1}^2 \sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2 \sigma_{a_3}^2} + \frac{q_{11}}{\sigma_{a_2}^2 \sigma_{a_3}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2 \sigma_{a_3}^2}} > 0.$$

Appendix 2: An expression of $\hat{\beta}^\#(0)$ in (26).

We consider the case with $K = 2$ and $p = 2$. From (20) and (23),

$$\mathbf{D}^*(\mathbf{0})^T = \begin{pmatrix} D_{11}(0), \dots, D_{N_1}(0), 1, 0 \\ D_{12}(0), \dots, D_{N_2}(0), 0, 1 \end{pmatrix} = \left(\mathbf{D}(\mathbf{0})^T, \begin{pmatrix} 1, 0 \\ 0, 1 \end{pmatrix} \right) \text{ and } V^{*-1} = \begin{pmatrix} \mathbf{V}^{-1}, \mathbf{0} \\ \mathbf{0}, \begin{pmatrix} \sigma_{a_1}^{-2}, 0 \\ 0, \sigma_{a_2}^{-2} \end{pmatrix} \end{pmatrix}. \quad (\text{A.14})$$

From (A.1) and (A.2),

$$\begin{aligned} Q &= \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \\ &= \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) + \begin{pmatrix} \sigma_{a_1}^{-2}, 0 \\ 0, \sigma_{a_2}^{-2} \end{pmatrix} \\ &= \begin{pmatrix} q_{11} + \sigma_{a_1}^{-2}, q_{12} \\ q_{12}, q_{22} + \sigma_{a_2}^{-2} \end{pmatrix} \end{aligned} \quad (\text{A.15})$$

From (A.1),

$$Q_1^{-1} = \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} = \frac{\begin{pmatrix} q_{22}, -q_{12} \\ -q_{12}, q_{11} \end{pmatrix}}{|Q_1|} \quad (\text{A.16})$$

From (A.3), $|Q| = |Q_1| + \frac{q_{11}}{\sigma_{a_2}^2} + \frac{q_{22}}{\sigma_{a_1}^2} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2}$. From (A.15) and (A.16),

$$\begin{aligned}
(\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}))^{-1} &= Q^{-1} \\
&= \frac{1}{|Q|} \left[\begin{pmatrix} q_{22}, -q_{12} \\ -q_{12}, q_{11} \end{pmatrix} + \begin{pmatrix} \sigma_{a_2}^{-2}, 0 \\ 0, \sigma_{a_1}^{-2} \end{pmatrix} \right] \\
&= \frac{1}{|Q|} \left[|Q_1| (\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}))^{-1} + \begin{pmatrix} \sigma_{a_2}^{-2}, 0 \\ 0, \sigma_{a_1}^{-2} \end{pmatrix} \right]
\end{aligned} \tag{A.17}$$

From (A.14),

$$\begin{aligned}
\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} &= \left(\mathbf{D}(\mathbf{0})^T, \begin{pmatrix} 1, 0 \\ 0, 1 \end{pmatrix} \right) \begin{pmatrix} \mathbf{V}^{-1}, \mathbf{0} \\ \mathbf{0}, \begin{pmatrix} \sigma_{a_1}^{-2}, 0 \\ 0, \sigma_{a_2}^{-2} \end{pmatrix} \end{pmatrix} \\
&= \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1}, \begin{pmatrix} \sigma_{a_1}^{-2}, 0 \\ 0, \sigma_{a_2}^{-2} \end{pmatrix} \right)
\end{aligned}$$

and from (15)

$$\begin{aligned}
\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{Y}^*(\mathbf{0}) &= \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1}, \begin{pmatrix} \sigma_{a_1}^{-2}, 0 \\ 0, \sigma_{a_2}^{-2} \end{pmatrix} \right) \begin{pmatrix} y_1(\mathbf{0}) \\ \dots \\ y_N(\mathbf{0}) \\ \beta_1^*(\mathbf{0}) \\ \beta_2^*(\mathbf{0}) \end{pmatrix} \\
&= \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}) + \Sigma_{\beta_p(\mathbf{0})}^{-1} \boldsymbol{\beta}^*(\mathbf{0})
\end{aligned} \tag{A.18}$$

Denote

$$\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}) = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}. \tag{A.19}$$

From (12), (24), (A. 16), (A.17), (A.18) and (A.19),

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^\#(\mathbf{0}) &= \left(\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \right)^{-1} \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{Y}^*(\mathbf{0}) \\
&= \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} + \begin{pmatrix} \sigma_{a_2}^{-2}, 0 \\ 0, \sigma_{a_1}^{-2} \end{pmatrix} \right] \left[\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}) + \boldsymbol{\Sigma}_{\beta_p(\mathbf{0})}^{-1} \boldsymbol{\beta}^*(\mathbf{0}) \right] \\
&= \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}) + |\mathcal{Q}_1| \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} \boldsymbol{\Sigma}_{\beta_p(\mathbf{0})}^{-1} \boldsymbol{\beta}^*(\mathbf{0}) \right. \\
&\quad \left. + \begin{pmatrix} \sigma_{a_2}^{-2}, 0 \\ 0, \sigma_{a_1}^{-2} \end{pmatrix} \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}) + \frac{\boldsymbol{\beta}^*(\mathbf{0})}{\sigma_{a_1}^2 \sigma_{a_2}^2} \right] \\
&= \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \hat{\boldsymbol{\beta}}(\mathbf{0}) + |\mathcal{Q}_1| \text{Cov}[\hat{\boldsymbol{\beta}}(\mathbf{0})] \boldsymbol{\Sigma}_{\beta_p(\mathbf{0})}^{-1} \boldsymbol{\beta}^*(\mathbf{0}) + \begin{pmatrix} \frac{g_1}{\sigma_{a_2}^2} \\ \frac{g_2}{\sigma_{a_1}^2} \end{pmatrix} + \frac{\boldsymbol{\beta}^*(\mathbf{0})}{\sigma_{a_1}^2 \sigma_{a_2}^2} \right] \\
&= \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \hat{\boldsymbol{\beta}}(\mathbf{0}) + [|\mathcal{Q}_1| \text{Cov}[\hat{\boldsymbol{\beta}}(\mathbf{0})] \boldsymbol{\Sigma}_{\beta_p(\mathbf{0})}^{-1} + \frac{1}{\sigma_{a_1}^2 \sigma_{a_2}^2}] \boldsymbol{\beta}^*(\mathbf{0}) + \begin{pmatrix} \frac{g_1}{\sigma_{a_2}^2} \\ \frac{g_2}{\sigma_{a_1}^2} \end{pmatrix} \right].
\end{aligned}
\tag{A.19}$$

Note that $1/|\mathcal{Q}_1|$ and $1/|\mathcal{Q}|$ are the generalized variances of $\hat{\boldsymbol{\beta}}(\mathbf{0})$ and $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$, respectively.

From (A.19), $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is a linear function of $\hat{\boldsymbol{\beta}}(\mathbf{0})$ and $\boldsymbol{\beta}^*(\mathbf{0})$ as well as their generalized variances, and variances and covariances of $\hat{\boldsymbol{\beta}}(\mathbf{0})$.

For a special case that $p = 1$, we have

$$\hat{\boldsymbol{\beta}}^\#(\mathbf{0}) = \frac{1}{|\mathcal{Q}|} \left[|\mathcal{Q}_1| \hat{\boldsymbol{\beta}}(\mathbf{0}) + |\mathcal{Q}_1| \frac{\beta_1^*(\mathbf{0})}{\sigma_{a_1}^2} \begin{pmatrix} \text{Var}[\hat{\beta}_1(\mathbf{0})] \\ \text{Cov}[\hat{\beta}_1(\mathbf{0}), \hat{\beta}_2(\mathbf{0})] \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{g_2}{\sigma_{a_1}^2} \end{pmatrix} \right]. \tag{A.20}$$

Table captions:

Table 1. Parametric optical critical dimension (OCD) fits to the data from in Figure 2 before and after the inclusion of data from an atomic force microscope (AFM).

Table 2. Parametric optical critical dimension (OCD) fits to the data in Figure 3 before and after the inclusion of data from atomic force microscopy (AFM). The five floating parameters are top width, middle width, bottom line width, line height, and the percent variation of the optical constant, n .

Table 1.

	OCD Fitting	AFM	OCD w/AFM
	$\hat{a}_k (\sigma_{\hat{a}_k})$	$a_k^* (\sigma_{a_k^*})$	$\hat{a}_k^\# (\sigma_{\hat{a}_k^\#})$
Top	119.68 (0.84)	119.21 (0.75)	121.27 (0.30)
Middle	111.32 (2.23)	117.32 (0.75)	116.03 (0.68)
Bottom	143.24 (1.32)	132.87 (0.75)	140.46 (0.44)

Table 2.

	OCD Fitting	AFM	OCD w/AFM
	$\hat{a}_k (\sigma_{\hat{a}_k})$	$a_k^* (\sigma_{a_k^*})$	$\hat{a}_k^\# (\sigma_{\hat{a}_k^\#})$
Top	33.7 (10.8)	37.6 (1.8)	39.2 (1.7)
Middle	48.9 (6.0)	44.7 (2.8)	50.1 (2.1)
Bottom	68.9 (8.3)	49.6 (5.9)	63.9 (3.1)
Height	60.0 (2.2)	55.5 (1.4)	58.4 (0.6)
n (% of nominal)	98.1 (1.0)		98.1 (0.5)

Figure captions:

Fig. 1. Cross-sectional view of a periodic line structure that serves as an input to the electromagnetic scattering simulations. The solid lines ending in diamonds are the height, h , and the pitch, p . The dotted lines ending in circles are the width of lines at the top, middle, and bottom. In the first example in Section 4, the height and pitch will remain fixed to reduce computational time, and the three line widths will be floated. Therefore, the vector a will have three components, but in general could have more or less.

Fig. 2. Examples of experimental data (markers) and library data fits (lines) for three line arrays from the overlay metrology advisory group (OMAG) 3 wafer. The four curves in each plot correspond to the four combinations of scan direction and orthogonal linear polarizations shown in the schematic.

Figure 3. An example set of experimental data (markers) and library data fits (lines) for the reflectivity from a patterned nitride line array on polysilicon.

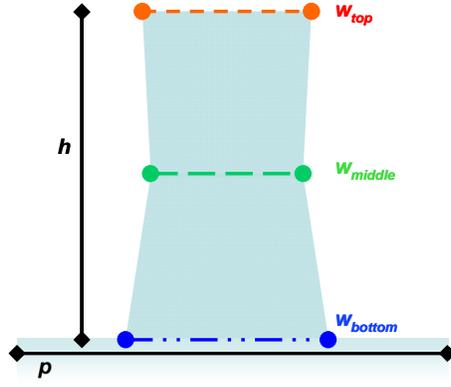


Figure 1.

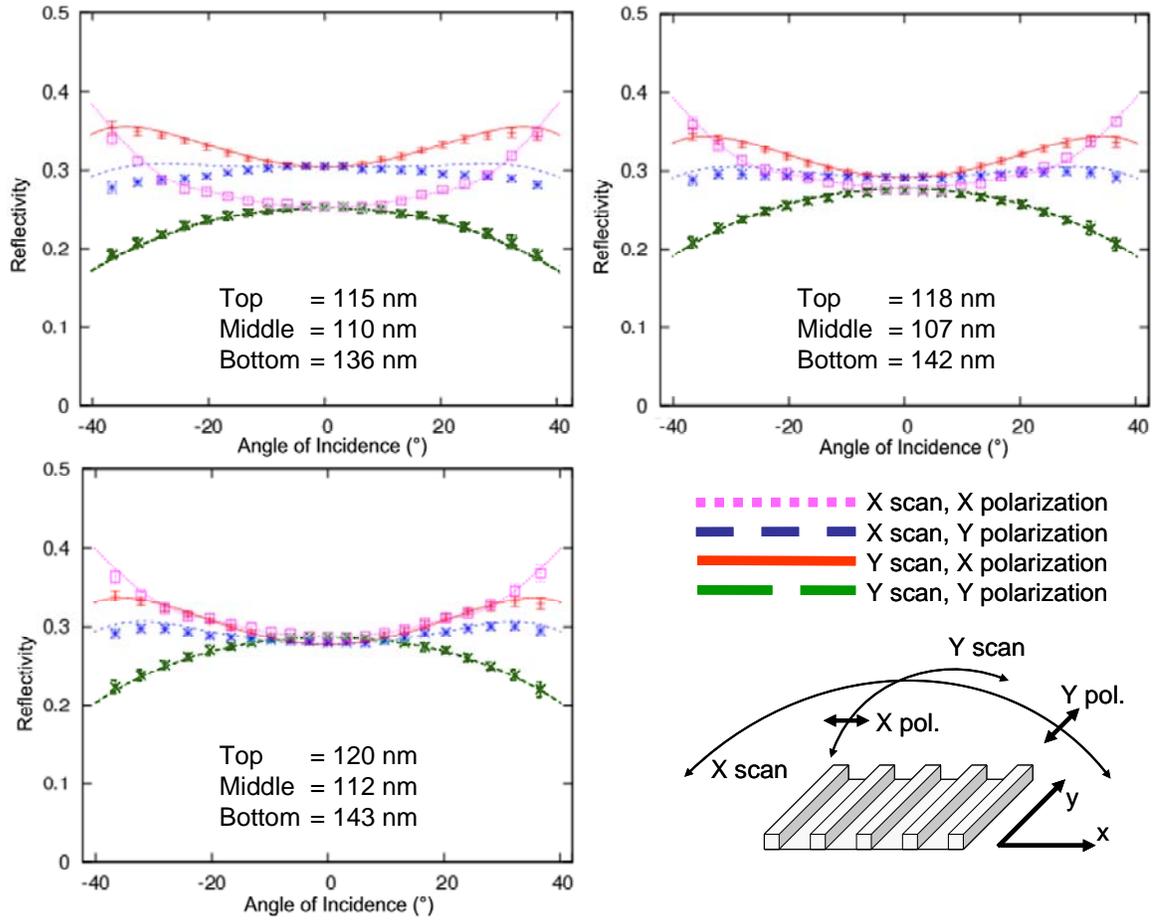


Figure 2.

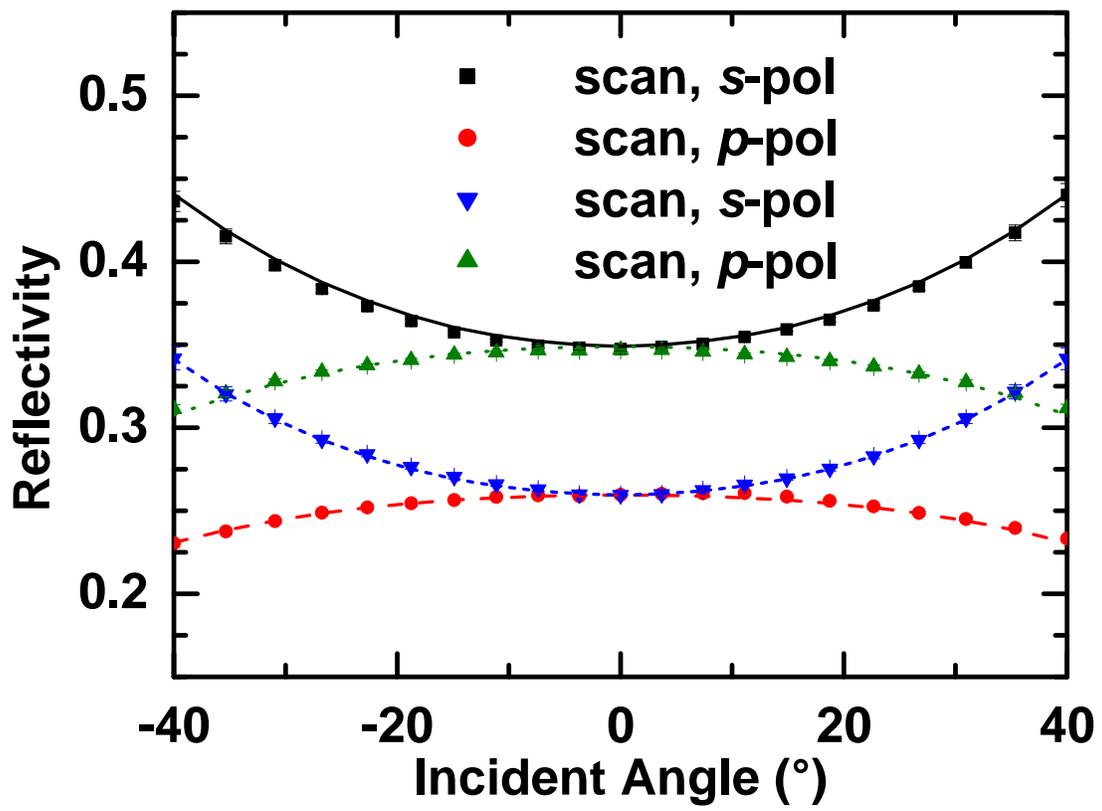


Figure 3.